SAMGrid Operations Policy

Adam Lyon (for the REX/Ops Group)

May 8, 2008

1 Introduction and definitions

This document describes the policies for supporting and operating the SAMGrid system at DØ. SAMGrid is defined to be the software and hardware infrastructure to run DØ production jobs on the Grid. Specifics are described below.

1.1 Software

The REX/Ops team will support the SAMGrid software components written by the SAMGrid development team. This software includes components running on the SAMGrid queueing nodes, forwarding nodes, SAMGrid native head nodes, station nodes, cache nodes, and durable location nodes. We support the JIM components, `sam_batch_adapter` components, and condor and globus components installed by the vdt. While we will not directly debug or fix components written by Condor, OSG and LCG, if we discover problems with those components we will lead the effort for a resolution with the appropriate party.

1.2 Hardware

We will directly support the queueing nodes, forwarding nodes, head nodes, station nodes, cache nodes, and durable location nodes that reside at Fermilab. Direct support means we will debug and resolve problems with those nodes be they hardware or software. We will lead the coordination with other parties (FEF, Fermigrid) to resolve hardware issues. We will support hardware at remote sites by installing SAMGrid software components there (if given log-in permission). We will also advise remote sites of hardware problems and standard node configuration. We are not directly responsible for hardware problems at remote sites (e.g. we will not arrange to fix cpus or buy new disks). We will, however, log in and install SAMGrid software on these nodes (but we expect the local sysadmins to install OS and the OSG or LCG components).

1.3 Scope

As mentioned above, REX/Ops takes complete first-line responsibility for the SAMGrid software components and nodes of the SAMGrid infrastructure at Fermilab. We will at the best effort level lead the coordination with remote personnel to resolve issues on nodes and components away from Fermilab. If we believe that problems at a remote site are affecting the system as a whole we will work with remote personnel to solve the problem. To that end, we expect to have contact information for personnel at each remote site (we do not have this information yet). We expect timely assistance in these matters. At our option, we may temporally resolve such problems by isolating or blocking problematic traffic from such sites.

1.4 Interaction with developers

We may from time to time require the assistance of the SAMGrid developers in investigating and resolving issues. We will ask for their help via the Issue Tracker and either they or us will record notes from their help into the issue tracker and our documentation. For the short term future, we will have weekly meetings where we review outstanding issues from the past week. If DØ plans a major increase in load, we may need assistance in evaluating usage and increasing system capabilities to scale up to the new load.

2 Operations level of service

While MC and primary production are extremely important, those activities are not as critical as the storage of online data from the detector to tape. Failure to store online data is the only emergency situation for REX/Ops. Expert shifters on call may be contacted at all hours for such problems. Problems in other components, including SAMGrid, do not warrant all hours contact of experts. The level of service for SAMGrid is described here and in the table below.

Time	**Support Level**	**Comments**	**Paging**
FNAL Business Hours	About one hour response	Expert shifter monitors issue tracker	Expert shifter as necessary.
Weekday Off Hours	Best effort	Expert shifter checks sam-oncall mailing list and IT once in evening	Expert shifter should not be called before 9am or after 9pm except for a data handling online emergency.
Weekends and FNAL holidays	Best effort	If possible, expert shifter checks sam-oncall list and IT at least once per day	Same as Weekday Off Hours

2.1 General practices

When an issue tracker ticket is assigned to the on call expert, (usually this action is done by the SAM shifter, or the expert themselves), the expert should post a comment to the ticket stating that he/she is investigating (that will send mail back to the ticket originator). The on call expert will investigate the problem, assign it to others if necessary, and eventually resolve the ticket when the problem is resolved. The on call expert is responsible for that ticket (regardless of other assignees) until their shift period is over or the ticket is closed. If the ticket is still unresolved when the new shifter comes on, it is passed to that new shifter. If the unresolved ticket is awaiting action from another assignee, then the active on call expert is responsible for making sure that action is completed.

If the on call expert is overloaded, he or she may transfer the ticket to the secondary on call expert. That secondary shifter then has responsibility for the ticket. The secondary shifter should acknowledge ownership by adding a comment to the ticket. If the ticket is still open when the secondary shifter shift is over, the ticket passes to the new primary shifter.

2.2 Response during FNAL business hours

During regular Fermilab business hours, the expert shifter on call will monitor the samoncall mailing list. Issue tracker tickets assigned to the on call expert by the regular SAM shifter will be acknowledged in a reasonable time (about one hour).

2.3 Response during weekday off hours

If possible, the on call shifter should check their mail and the Issue Tracker sometime in the evening and attempt to handle problems that have emerged. Note that most likely problem will not be resolved until the next business day, but a best effort attempt should be made. Also note that most of the hardware at Fermilab supporting SAMGrid is **not** supported 24/7 by system administrators.

Since SAMGrid is not a detector critical system, the on call expert should not be called or paged before 9am or after 9pm for SAMGrid problems. Users seeing a problem during this time should make a new SAM issue tracker ticket. The offline shifter may call or page the on call expert at the next appropriate time.

2.4 Response during weekends and FNAL holidays

Response during this time is best effort. The on call expert should check e-mail and the Issue Tracker at least once per day and respond to problems if possible. If the on

call expert will not have internet access, he or she should make arrangements with the secondary shifter to do this work. Rules for off hours weekday support apply here.

3 A process used to triage and assign tickets

The process for handling issue tracker tickets is quite simple.

- Anyone may submit a new issue to the issue tracker by sending e-mail to `d0sam-admin@fnal.gov` or using the issue tracker web interface. All problem reports about SAMGrid must proceed through the issue tracker. Those sending mail about problems directly to shifters will be told to resend their message into the issue tracker.
- Problems noticed by the Grid Operations Center (GOC) should be entered into the issue tracker system by the expert shifter (more on this below).
- If the regular offline shifter cannot solve the problem, he or she escalates the issue to the on call expert. The expert shifter can also notice issues in the issue tracker and escalate them if he or she knows it is an expert issue. (The expert shifter here is the primary shifter unless the primary has asked the secondary for help in addressing new issues).
- The expert shifter then immediately acknowledges that they own the ticket.
- The expert shifter works the problem, assigns the ticket to others if necessary, and ultimately ensures that the ticket is resolved.
- If the expert shifter ends his or her shift and there are still open tickets, those tickets transfer to the incoming expert shifter.

3.1 GOC tickets

The OSG Grid Operations Center will send issues regarding the DØ VO to the `DZEROVO-SUPPORT` mailing list. Adam Lyon, Robert Illingworth, and Steve Sherwood are the REX/Ops personnel subscribed to that list. If a GOC ticket arrives, we will take appropriate action by either entering it into the SAM issue tracker or passing the ticket to the appropriate party.

4 Deployments and changes to the production system

Now that the transition of operations to REX/Ops is complete, the developers should no longer access the production systems except for troubleshooting and monitoring. Development for new components and bug fixes should occur on designated development and test nodes (see below), not production. Deployments on the production system will be made by the REX/Ops group, with advice and assistance from the developers if necessary. The developers should not change any item on the production system without permission form the SAMGrid project manager.

4.1 Deployment of new code components

In the complicated distributed environment of SAMGrid, full scale production level testing is nearly impossible. There is no test farm that can mimic all of the environments and variables that the production system faces. Therefore, some failures of new code are, unfortunately, unavoidable. However, we will make a best effort to catch obvious problems before new code is deployed by testing code on a small scale test system (see test system below).

4.2 Deployment of new configuration

Management of the SAMGrid configuration is very complicated and prone to mistakes. We are developing a system where most configuration files are in CVS. The goal is to have a system where a change to the configuration is made in CVS, and that change is pushed to all of the appropriate nodes.

5 Test system

REX/Ops will maintain several nodes that parallel main components of the production system - namely the queueing and forwarding nodes. We will start with an OSG forwarding test node and will eventually add an LCG forwarding test node. Since we do not have a test farm, these test nodes will submit small scale jobs to real production farms for testing. When there are no new releases to test, the nodes will perform continuous tests of the current release. New releases and configurations will be deployed to the test system and tried. If there are no problems, then the new release/configuration is deemed production ready and deployed into production. If a downtime is needed we will coordinate with the DØ offline management.

We may for a short time loan out the test system to the developers for special short-term development projects. The developers must return the nodes to their previous state when they are finished.

5.1 Emergency fixes

If the production system is in an inoperable state due to a failure of the SAMGrid code or configuration, the SAMGrid project manager in consultation with experiment offline managers may opt to deploy fixes directly to production and the test nodes and then do tests along with the production jobs. For such a situation, the priority is to get production working as quickly as possible. The end users will be warned that there may be further bugs or problems that may cause their jobs to fail and necessitate further releases.

If the system is in an operable state, then new deployments must proceed first through the test system.

6 System failures

Very little of the SAMGrid system is redundant, though some components are duplicated to share the load. Losing a node without a corresponding drop in load will most likely lead an overloaded system. The response to a hardware failure is made in conjunction with experiment management. If the node cannot be replaced or fixed quickly and its share of the load is necessary to continue, then we may convert a test node into a production node. It may take time (a day or two) for this conversion to occur since we may have to install software and certificates. Furthermore, the test systems will tend to be old and underpowered and therefore they will not totally replace the lost node. The experiment will have to adjust the load accordingly. When the broken node is replaced or repaired, we expect the return of the test node.

7 Keeping the system up to date

We expect the SAMGrid developers to take reasonable steps to keep the third party software (VDT, Condor, etc) as up to date as possible for the lifetime of SAMGrid. A schedule and detailed plan for such updates needs to be worked out.

8 Summary

The REX/Ops group will support the SAMGrid system as described above. Off hours support is on a best effort basis. We have a system for triaging issues from users and the OSG Grid Operations Center. A small scale test system will verify basic operation of new releases before deployment to production.